#So Berlin?

Clustering Germany's biggest cities according to their nightlife, average income and rental prices

Felix Klapprott

Mai, 2020

Part II: Data

I decided to focus the analyses on cities with more than 500 000 Inhabitants. A useful overview is found freely available on the German Wikipedia at https://de.wikipedia.org/wiki/Liste_der_Gro%C3%9F-_und_Mittelst%C3%A4dte_in_Deutschland . An extracted list will be the basis of all analyses as it provides a list of official names and the needed selection of all German cities with more than 500.000. The list will be used to combine official statistics, geo-data and foursquare data.

Additionally rental prices for these cities are needed. The data for the calculation of rents in the fourteen German cities with more than 500,000 inhabitants were 234,000 offers advertised on the online platform immowelt.de. Only those offers that were in increased demand were taken into account. The prices reflect the median of the rental apartments and houses offered in the first half 2019. The rental prices reflect the median of the net cold rents for new rentals. The data set (German) including a comparison to 2009 is accessible (premium/university access) at https://de.statista.com/statistik/daten/studie/167163/umfrage/mietentwicklung-in-den-deutschen-grossstaedten/ . As it is not publicly available I will download it and convert to csv in order to work with it in Python. Rental prices will be used as a variable in the later cluster analysis tofind patterns among Germanys biggest cities.

The most recent data I could find on accessible income per capita in German cities was collected in 2016 by official offices using questionnaires. The data set can be found on https://de.statista.com/statistik/daten/studie/998971/umfrage/verfuegbares-einkommen-in-den-groessten-staedten-in-deutschland/ . As it is not easily available I will download it too and convert to csv in order to work with it in Python. Average Income will be used as a variable in the later cluster analysis tofind patterns among Germanys biggest cities.

The foursquare API will provide me with useful information about the type of venues in a given city. I will focus only on venues which are registered as "Nightlife spot". I will calculate relative frequencies of different venue types per city to find similarities between cities, i.e. for example distinguishing cities where you would find mainly cocktail and wine bars as opposed to night clubs, pubs and beer gardens.